BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2014-25

Improving Overnight Loan Identification in Payments Systems

by Mark Rempel

www.bank-banque-canada.ca

Bank of Canada Working Paper 2014-25

June 2014

Improving Overnight Loan Identification in Payments Systems

by

Mark Rempel

Funds Management and Banking Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
a.mark.rempel@gmail.com

Bank of Canada working papers are theoretical or empirical works-in-progress on subjects in economics and finance. The views expressed in this paper are those of the author. No responsibility for them should be attributed to the Bank of Canada.

ISSN 1701-9397

© 2014 Bank of Canada

Acknowledgements

I thank the Canadian Payments Association for access to their data and feedback. I am also grateful for comments and advice from Jason Allen, Narayan Bulusu, James Chapman, Jonathan Chiu, Scott Hendry, Kim P. Huynh, Vathy Kamulete, Miguel Molico, Lukasz Pomorski, Francisco Rivadeneyra and Tsz-Nga Wong, as well as seminar participants at the Bank of Canada and Canadian Economic Association. Special thanks to Boyan Bejanov for technical assistance with EDITH, the Bank of Canada's computing cluster. Any errors are my own.

Abstract

Information on the allocation and pricing of over-the-counter (OTC) markets is scarce. Furfine (1999) pioneered an algorithm that provides transaction-level data on the OTC interbank lending market. The veracity of the data identified, however, is not well established. Using permutation methods, I estimate an upper bound on the daily false positive rate of this algorithm to be slightly above 10%. I propose refinements that reduce the bound below 10% with negligible power loss. The results suggest that the inferred prices and quantities of overnight loans do provide viable estimates of interbank lending market activity.

JEL classification: E42, E44, C53, C38, G10
Bank classification: Econometric and statistical methods; Financial markets; Interest rates; Payment clearing and settlement systems

Résumé

Il y a peu d'information au sujet de la répartition des actifs et de la formation des prix sur les marchés de gré à gré. En 1999, Furfine a conçu un nouvel algorithme qui fournit des données relativement aux transactions effectuées sur le marché du financement interbancaire. La validité de ces données n'est cependant pas bien établie. En procédant à une permutation des données, j'estime que la limite supérieure du taux de faux positifs quotidien généré par cet algorithme est légèrement au-dessus de 10 %. Je propose des améliorations permettant de réduire cette limite sous 10 % tout en subissant une perte de puissance statistique négligeable. Les résultats donnent à penser que les quantités et les prix des prêts à un jour déduits grâce à cette méthode constituent des estimations valables de l'activité sur le marché du financement interbancaire.

Classification JEL : E42, E44, C53, C38, G10
Classification de la Banque : Méthodes économétriques et statistiques; Marchés financiers; Taux d'intérêt; Systèmes de paiement, de compensation et de règlement

1 Introduction

Frictions in over-the-counter (OTC) markets played a significant role in the 2007/2008 financial crisis. Study of these markets is limited by the scarcity of data on the formation of prices and quantities. Information on one OTC market, the interbank lending market, has been notably leveraged by Ashcraft and Duffie (2007) to understand the dynamics of OTC markets in general.

The interbank lending market is itself important to study as the primary conduit for monetary policy and an integral source of intraday liquidity. Frictions in this market can have significant spillovers into other markets, as Acharya and Merrouche (2013) show on the volumes and rates of both household and corporate lending in the United Kingdom.

The empirical identification of such frictions relies upon transaction-level data. While weighted-average rates of overnight loans can be obtained via surveys (e.g., effective federal funds rate, or Canada's overnight money market financing (OMMF) rate), the OTC nature of the interbank lending market implies that very little information on the volume of loans is directly available. Information on the effective rates is important as monetary policy is generally implemented through a target of that rate, and decisions on intervention are predicated on observed deviations from the target. However, information on the quantities is also crucial, since during the crisis the supply of federal funds remained fixed while demand increased markedly (Afonso, Kovner, and Schoar (2011)).

The transaction-level data typically used is acquired by applying Furfine's (1999) algorithm to a payments system like Fedwire. This algorithm has been refined as in the work of Demiralp, Preslopsky, and Whitesell (2006) and Hendry and Kamhi (2009), as well as generalized to term loans by Kuo et al. (2013), but remains effectively a collection of filters to identify interday payment pairs that reflect the observable characteristics of an interbank loan and repayment.

This paper assesses the susceptibility of Furfine's (1999) algorithm to spurious matching and proposes some means to improve its accuracy. Until recently, little analysis on the validity of the algorithm (hereafter referred to simply as the 'Furfine algorithm') had been conducted. Armantier and Copeland (2012) acquired Fedwire transaction identifiers from two large participants and compared the list of identified federal funds loans directly between the two banks against the output from the Furfine algorithm. While their estimates of false positives and negatives are high, they recognize that the estimates may not be representative of other participants and may simply be explained by correspondent banking.

I conduct a permutation-style test to estimate the rate of spurious matches generated by the Furfine algorithm. I do so by re-pairing non-adjacent days of payments together, applying the Furfine algorithm to them as if they were adjacent and comparing the spuriously generated quantities against the quantities obtained on the correct ordering of days. I impose structure on these quantities drawn from the work of Storey (2002), Storey and Tibshirani (2001) and Efron (2004) on positive false discovery rates, in order to ascribe this ratio of quantities as a conservative estimate of the upper bound of type I error.

From a complete sample of payments between January 2005 to December 2011, I estimate this upper bound at 12.5% with 99% confidence. Further, I characterize the obtained false positives based on observables and propose a refinement involving k-Nearest Neighbours that leverages this information. I find that this refinement reduces the upper bound below 10% (8.3%/9.6% at the 95%/99% level) without the levels of implied power loss of other studied variants. Finally, I assess the algorithm's exposure to foreign exchange transactions as a source of type I error and propose a heuristic method to mitigate its influence.

The remainder of the paper is organized as follows. Section 2 provides a brief overview of real-time gross settlement systems and the application of the Furfine algorithm to such payments systems. Section 3 assesses the efficacy of the Furfine algorithm and characterizes the observable drivers of the false positive rate. I propose a refinement in Section 4 and compare its efficacy against other known refinements. In Section 5, I consider other sources of type I error not accounted for in my primary methodology. I explore the economic significance of my refined algorithm and demonstrate its use for policy-making in Section 6, and offer some concluding thoughts in Section 7.

2 Background

Since the turn of the century, real-time gross settlement (RTGS) systems have become the norm amongst the more-developed economies. As the name suggests, these electronic, large-value payments systems allow for real-time finality and settlement of payments between participating institutions. Federally administered by a nation's central bank, these systems effectively eliminate intraday settlement and credit risk. While transfers between clients and subsidiaries of the RTGS participants will not enter into the system, most other payments between financial institutions will generally be transacted within the RTGS.

Seeking to capture the set of interbank loans from this universe of transactions, Furfine (1999) pioneered an algorithm that identified interday bilateral payment flows which resembled interbank overnight loans. The algorithm is essentially a collection of filters, requiring that (i) the principal payment be a certain size and quoted in rounded increments, (ii) the principal and repayment legs be on adjacent business days and between the same counterparties (with reversed payment flow across the two days), and (iii) the implied interest rate from the repayment observed the following day fall within a 'reasonable' range of the overnight target.

Various adaptations of this algorithm have been proposed. For instance, Demiralp, Preslopsky, and Whitesell (2006) proposed the requirement that the inferred interest rate be quoted to market standards (i.e., in the United States, that the interest rate be in whole basis points or 1/32nds of a percentage point), and Hendry and Kamhi (2009) attempted to filter out foreign exchange transactions.

3 Efficacy of the Furfine Algorithm

3.1 Overview

The collection of filters that are the 'Furfine algorithm' seek to match the necessary conditions of an overnight loan. The general validity of the output from the algorithm, however, is not well established. Recently, Armantier and Copeland (2012) assessed the bank-to-bank level of identification of the Furfine algorithm by acquiring a list of payments identified to be overnight loans by two active banks in the federal funds market. Comparing this external reference list with the payments identified by the algorithm in the first quarter of 2007 and 2011, respectively, they find that, on average, 81% of the algorithm's output between those two counterparties was not in the reference list.

It is unclear whether the incompatibility between the reference list and the algorithm's output arises from spurious matching or through non-RTGS participant banks trading loans through direct clearers. While certainly a limitation for some questions, for many research and policy-making purposes near-perfect accuracy in identifying the counterparties of an interbank loan is less important than the correct estimation of the distribution of rates and volumes.

I seek in the remainder of this section to assess the quantity of the daily output of the Furfine algorithm that is spuriously generated. To do so, I essentially apply the Furfine algorithm to randomly paired, non-adjacent days (which cannot possibly have any true overnight loans). I

compare the quantity of loans identified by these random day pairings against those identified by the standard running of the algorithm to estimate the daily rate of spurious matches.

3.2 Statistical Methodology

Given the null, $H_{0i}^{dd'}$, that a given payment, x_i^d, is not the principal transfer of an overnight loan between day d and d', the Furfine algorithm can be interpreted as a sequence of tests, with test statistics $(T_i^{dd'})_{i=1}^m$, where

$$T_i^{dd'} = min\left\{|\text{interest rate}(x_i^d, x_j^{d'}) - \text{ON target}| : sender(x_j^{d'}) = receiver(x_i^d), sender(x_i^d) = receiver(x_j^{d'})\right\},$$

where ON target is the overnight target rate. Each test statistic T_i gives the closest possible inferred interest rate to the target of payment x_i on day d from the set of candidate payments on day d'.

The associated rejection region, Γ_λ is the set of payments for which the difference in the implied interest rate from target falls below a threshold, λ (typically set at 50 basis points), and that are classified as candidate overnight loans by the algorithm. Specifically, $\Gamma_\lambda \equiv \{T_i : T_i \leq \lambda\}$ and the daily number of loans identified by the Furfine algorithm between day d and d' is $S_{dd'} \equiv \sum_i^m \mathbf{1}\{T_i^{dd'} \in \Gamma_\lambda\}$.

The proportion of S that is type I error, referred to as the false discovery rate, is

$$pFDR \equiv \frac{\sum_i^m \mathbf{1}\{T_i^{dd'} \in \Gamma_\lambda, H_{0i}^{dd'}\}}{S_{dd'}}.$$

Taking the numerator as not directly observable, I obtain an estimate of this quantity by applying the algorithm to non-chronological days following a similar daily payment distribution.[1] Specifically, to estimate the false discovery rate for a given day d, I sample (with replacement) B^R payment days to be treated as the repayment day for d (d_r), provided d_r and d share the same overnight policy rates, average the number of payments falsely identified as the principal of an overnight loan ($E[S^0]_d = \frac{1}{B^R}\sum_{d'}^{B^R} \mathbf{1}(S_{dd'}^0 > 0)$ and inflate the estimate by the proportion of resampled pairs yielding no false positives ($Pr(S_d^0 > 0)$).

I acquire a distribution of these estimates by randomly sampling principal payment days (d)

[1] I assume here that payment days sharing the same target interest rate are drawn from the same payment distribution and that payments are weakly dependent on each other.

B^N times $(d^1, ... d^N)$, yielding:[2]

$$\left(\widehat{pFDR}_{d_s}\right)_{d_s=d^1}^{d^N} = \left(\frac{E[S^0]_{d_s}}{Pr(S^0_{d_s}>0)S_{d_s d'_s}}\right)_{d_s=d^1}^{d^N}.$$

To illustrate, suppose I wish to estimate the false discovery rate of overnight loans traded on only one day (d). I sample day $d-2$ to replace day $d+1$ as the repayment day for which there cannot be any overnight loans by definition and which therefore is spurious. For simplicity, set $B^R = 1$, so that this is the only repayment day sampled. If there is at least one loan identified when I apply the algorithm on day d, $d-2$, then $Pr(S^0_{dd'} > 0) = 1$ and the measure of the false discovery rate simplifies to the number of loans identified on day d with return day $d-2$ against the number of loans that I originally obtained from day d and $d+1$. In this case, the $pFDR$ is the percentage of spurious matches out of the total originally matched. With $B^R > 1$, the average number of spurious matches replaces the numerator above, yielding the average percent spurious out of the total for overnight loans identified as initially traded on day d.

Asymptotic and finite-sample support for the 'conservativeness' of these estimates is provided by Storey (2002). Storey and Tibshirani (2001) show how this support extends to various forms of dependence in hypotheses and test statistics.[3] Finally, the time-series resampling under weak dependence amongst the payments data draws support from the serial testing methodology developed by Chu et al. (2013).

3.3 Estimated False Positive Rate Bounds

I apply the above methodology to Canada's RTGS-equivalent payments system, the Large Value Transfer System (LVTS), on the full sample of payments between 3 January 2005 and 31 December 2011 (excluding holidays and the days preceding them).[4,5] Setting $B^N = 1{,}500$ and $B^R = 20$, I obtain 30,000 false principal, return payment day pairs, culminating in $B = 1{,}500$ estimates of the $pFDR$.

Figure 1 depicts the empirical distribution of the upper bound on the daily false discovery

[2]Formally, the inclusion of the inflation term $Pr(S^0_d > 0)$ provides an estimate of the positive false discovery rate, not the false discovery rate. See Storey (2002) for motivation on the importance of utilizing the pFDR rather than the false discovery rate. Given the near-perfect probability of identifying at least one overnight loan a day, the distinction is not critical.

[3]In particular, their study of weak dependence (or 'loose dependence') amongst the test statistics and independence amongst the null statistics maps closest to the context of the Furfine algorithm.

[4]In addition to excluding all federal and provincial holidays, federal U.S.-specific holidays are also excluded, due to the distinct payment distribution observed on these days.

[5]For details on the LVTS and how it relates to other RTGS systems, see Arjani and McVanel (2006).

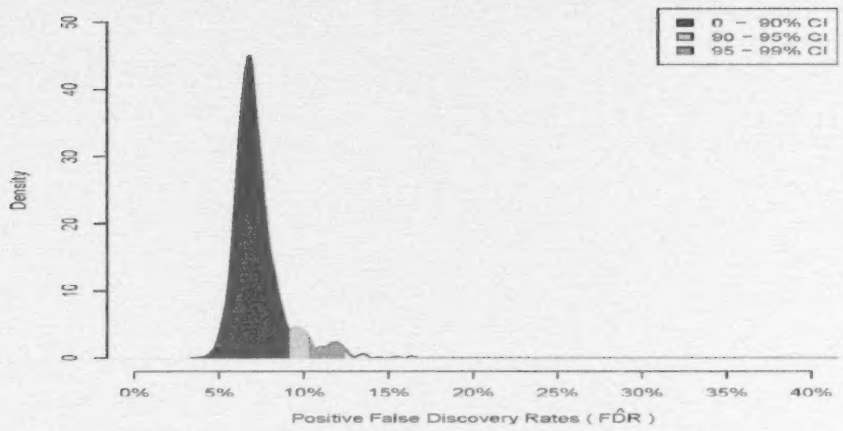


Figure 1: Empirical Distribution of False Discovery Rates, 2005Q1 - 2011Q4

rate. The shaded regions depict the 90%, 95% and 99% (from darkest to lightest) confidence bounds on the pFDR, which are at 9.1%, 10.3% and 12.5%, respectively. This suggests that for over 99% of the standard days in the payment cycle, less than 15% of the output is generated by spurious matches. On the downside, the upper bound on this error rate is below the 5% level less than 1% of the time.

Since this sample contains distinct interest rate and payment flow regimes, I also conduct the same estimation procedure on five subsamples (determined by the Ploberger and Krämer (1992) OLS-CUSUM test on the number of overnight loans identified using the significance boundaries proposed by Zeileis (2000)): two pre-crisis samples, 2005Q1 - 2006Q1 and 2006Q2 - 8 August 2007 (BNP Paribas' asset freeze); the financial crisis, 9 August 2007 to 21 April 2009 - the start of Canada's effective lower-bound period; the effective lower-bound period, 21 April 2009 to 31 May 2010; and the post-crisis period, 1 June 2010 to 30 December 2011. The estimated bounds for these samples are in rows 2-6 of Table 1.

In general, the full-sample results found are robust to the various subsamples estimated, with the average estimated upper bound on the false positive rate below 10% in all cases. However, the bounds on the estimated rate of false positives vary considerably across the subsamples. They are substantially higher for the early pre-crisis and crisis periods, with the 95% confidence intervals for both samples containing 10%, while the other samples are strictly below 10% at the same significance level.

		$\Pr(\widehat{pFDR} > x)$					
	Mean($\widehat{pFDR}$)	$x = 0.2$	$x = 0.1$	$x = 0.05$	$x = 0.01$	B	% Full Sample
Full Sample	7.32	0.07	6.13	99.13	100.00	1500	100.00
2005Q1 - 2006Q1	9.20	0.00	34.00	98.20	100.00	1500	20.23
2006Q2 - BNP Paribas	6.87	0.00	1.07	96.80	100.00	1500	19.73
BNP Paribas - ELB	7.19	0.20	5.47	96.80	100.00	1500	25.26
Effective lower-bound (ELB)	6.84	0.00	0.47	99.00	100.00	1500	13.28
ELB - 2011Q4	6.63	0.00	0.27	98.67	100.00	1500	21.50

Table 1: Estimated False Discovery Rates (in %) using permuted/resampled payments data within subsamples.
Estimated upper bounds on the daily type I error rate in percentage terms. Columns 2-5 provide the estimated probability of daily type I error rates above 20%, 10%, 5% and 1% of the algorithms output, respectively. The last column provides the number of loans identified in each subperiod relative to the full sample.

4 Improving Algorithmic Identification

4.1 Motivation

Numerous attempts have been made to improve the accuracy of interbank loan identification, but without any formal demonstration of improvement. For instance, Demiralp, Preslopsky, and Whitesell (2006) and Hendry and Kamhi (2009) propose rounded interest rate filters to reflect market quoting conventions, while Ashcraft and Duffie (2007) and Allen, Hortaçsu, and Kastl (2011) restrict their samples to the end of day when the reserve positions of the banks become evident.

In this section, I propose a novel treatment of non-uniquely identified payment pairs. Table 2 suggests the importance of this treatment where a disproportionate share of identified false positives arises from non-uniquely identified payment pairs. Intuitively, if there exist systematic, non-loan payment patterns with defined relationships across days, the potential exists that entire sets of non-uniquely identified payment pairs are in fact spuriously generated.

Quantiles	1%	25%	50%	Mean	75%	99%
Original sample	7.69	15.15	17.31	17.61	19.76	28.23
Permuted sample	51.53	57.74	60.52	61.19	63.76	77.26

Table 2: Percentage of Furfine identified loans with non-uniquely paired payment legs.
The first row shows selected quantiles of the percentage of non-uniquely identified loans from all those identified by the original application of the Furfine algorithm within a day. The second row shows the same measurements for the permuted sample (i.e., the Furfine algorithm run on non-adjacent days) as described in Section 3.

Hendry and Kamhi (2009) argue that one potential cause of non-unique pairs is foreign exchange transactions between the domestic currency (CAD in their case) and the U.S. dollar

or euro. High volumes of independent foreign exchange (FX) transactions on day d and d' with sufficient exchange rate variation may lead to pairs of FX transactions with inferred interest rates that satisfy the filters of the algorithm.[6] In such instances, it is likely that multiple payments will be spuriously matched with the same principal or return payment.

The other, more common concern with non-uniquely identified loans is the correct selection from non-uniquely identified payment pairs. Selection mechanisms are somewhat ad hoc. Armantier and Copeland (2012) use the payment pair with the median interest rate amongst the set, Hendry and Kamhi (2009) use the closest to the overnight target, and Demiralp, Preslopsky, and Whitesell (2006) use the term of uniquely paired loans to select the closest match amongst the non-unique set.

4.2 Proposed Refinement

My proposed refinement is predicated on the assumption that uniquely identified payment pairs are more likely overnight loans than non-uniquely identified pairs. I compare the implied loan characteristics of the uniquely identified 'loans' against those from the non-uniquely identified set and determine which, if any, of the non-unique set are sufficiently similar to the unique set. My proposed refinement offers two material improvements to past implementations of the algorithm. Where other implementations attempt to select the 'best' non-unique payment pair from the set, this method first determines whether there is in fact a reasonable potential overnight loan in the set and then selects the best pair. This allows for potential improvements in the rate of false positives.

The exact mechanism is an application of the k-Nearest Neighbours (k-NN) algorithm.[7] The k-NN algorithm is a computationally simple non-parametric method useful in classification exercises, where one has two sets of covariates X_0/X_1 and classifiers Y_0/Y_1, with all but Y_1 observed. The pair (X_0, Y_0) is used as a training set to offer predictions on the associated classifier (Y_1) with covariates X_1. The algorithm forms these predictions of Y_1 based on the majority vote of the 'k' closest observations in the training set X_0 to the corresponding X_1.

For the training set, I use the loan characteristics of the uniquely identified loans along with synthetically generated non-loan payment pairs (hereafter referred to as 'faux loans'). These synthetic 'faux loans' are randomly paired payments between the same two participants which,

[6] See Section 5 for further discussion of the exposure of the Furfine algorithm to foreign exchange transactions.
[7] See Hastie, Tibshirani, and Friedman (2009) for further details on this method.

unlike the Furfine output, are not constrained to be chronological in payment/repayment dates, and do not satisfy the interest rate filter.[8] To keep the density of the non-loan payment and uniquely identified loan set equivalent, I generate a 'faux loan' for each unique loan identified between two RTGS participants.

I compare the observable characteristics of this training set - the time the principal is submitted into the RTGS, the value of the principal, the duration of the loan, the spread from target and indicator of whether the implied interest rate is rounded to the nearest half-basis point - against the set of non-uniquely identified payment pairs, or prediction set, using the 'Mahalanobis' metric and a k of 15.[9] Non-uniquely identified payment pairs that have at least half of their k-Nearest Neighbours uniquely identified as loans are admitted as candidate overnight loans, while the others are discarded.[10]

This approach can be considered as a variant of end-of-day and rounded basis point refinements, just applied to the non-uniquely identified payment set and with softer univariate cut-off boundaries. Figure 2 contrasts the intraday distribution of the number of loans and values for the original Furfine output (top left), unique loans (top right), faux loans (bottom left) and non-unique Furfine loans (bottom right) in the last third of 2011.[11]

4.3 Performance

To properly evaluate the efficacy of my proposed refinement, careful attention must be given to the data used as training data for the resampled days. Independently sampling pairs of days for the training data set and the evaluation set, I use the characteristics of the loans uniquely identified by the algorithm in the first stage with principal payment on the corresponding sampled day as a proxy for 'true loans' and characteristics of randomly paired payments from the resampled days as a proxy for the counterfactual ('faux' loans), for the training set. To limit the likelihood of using the same payments for both the training and evaluation sets, I do not require the training and evaluation sets to share the same target rate.

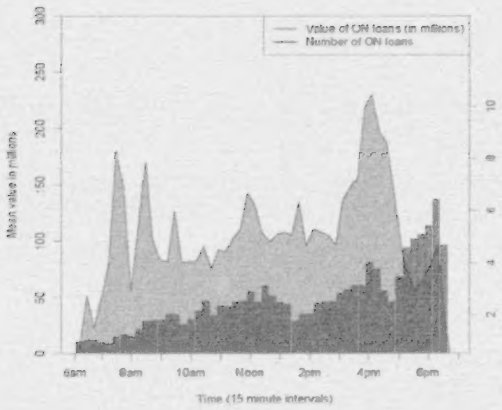
This should provide a conservative estimate of the refinement's influence on the false positive rate, since the training sample paired with the evaluation could come from any point in the

[8] For random pairs with negative inferred interest rates, I take the absolute value of the rate, effectively reversing the payment flow.

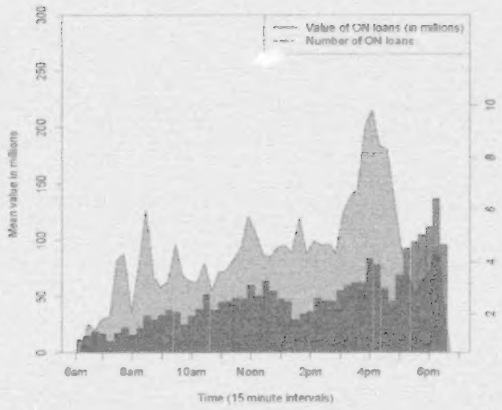
[9] Optimal k determined by cross-validation on a subsample of payments between 2005 and 2011.

[10] If multiple candidate payment pairs still remain involving the same payment leg, the pair with the interest rate closest to the target is chosen.

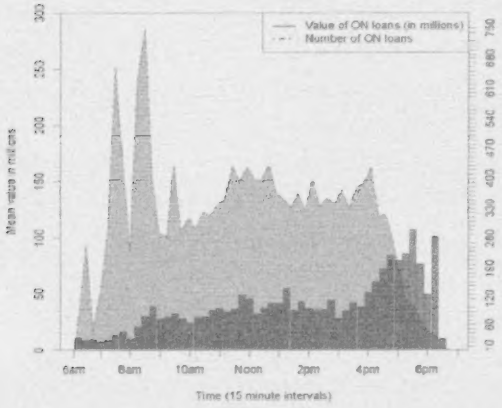
[11] The sample period selected is arbitrary and used for illustration only. The discussed inferences do not qualitatively differ with other subsamples.



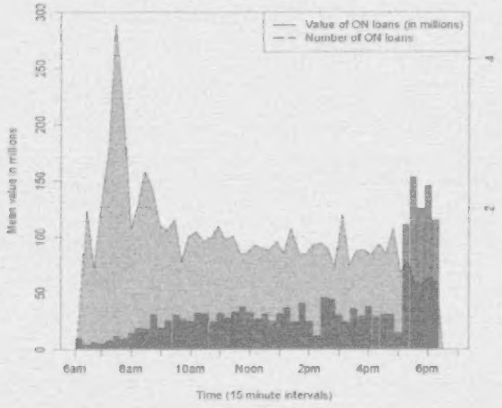
(a) Furfine-identified overnight loans, 1 Sept. 2011 - 30 Dec. 2011.



(b) Uniquely identified Furfine loans, 1 Sept. 2011 - 30 Dec. 2011.



(c) Randomly paired payments ('faux loans'), 1 Sept. 2011 - 30 Dec. 2011.



(d) Non-uniquely identified Furfine loans, 1 Sept. 2011 - 30 Dec. 2011.

Figure 2: Joint time/value distributions of the training data for 1 Sept. 2011 - 30 December 2011.

sample, whereas one could partition the data into different payments sets for the same payment or interest rate regime. In particular, utilizing smaller rolling samples for training and prediction increases the computational burden significantly but the accuracy as well.

The estimated upper bound on the false positive rate for my proposed refinement is given in row 1 of Table 3. Comparing against the baseline algorithm (row 2), I achieve a 15% improvement in the average false positive rate. Even more striking is the tightened 99% confidence interval on the daily false positive rate, which falls below the 10% significance level (nearly a 20% reduction in the 99% bound vis-à-vis the baseline).

Comparing these bounds against those from other refinements found in the literature suggests that the other refinements do not have lower upper bounds on false positive rates than the baseline algorithm, let alone my proposed variant.[12] Furthermore, my refinement outperforms the other refinements in the estimated rate of false positives while simultaneously limiting the exposure to type II error. This is demonstrated in the last column of Table 3, measuring the percentage of the overnight loans identified by the refinement against the baseline Furfine algorithm, where my refinement reduces the number of overnight loans identified relative to the baseline algorithm by only 8.37%, while the exact rounded interest rate requirement drops the sample by 87.76%.

		$\Pr(\widehat{pFDR} > x)$					
	Mean($\widehat{pFDR}$)	$x = 0.2$	$x = 0.1$	$x = 0.05$	$x = 0.01$	B	% Baseline
k-NN refinement	6.18	0.07	0.73	88.73	100.00	1500	91.63
Baseline algorithm	7.32	0.07	6.13	99.13	100.00	1500	100.00
End of day	17.64	33.27	34.00	79.73	100.00	1500	19.39
Exact rounded interest	18.59	34.00	91.33	99.93	99.93	1500	12.24
Rounded interest	41.17	91.93	99.40	99.93	99.93	1500	12.80

Table 3: Estimated False Discovery Rates upper bounds for Furfine Variants.
The estimated false discovery rates for each refinement are obtained in the same fashion as the baseline, with the exception being that the denominator is no longer the number of overnight loans identified by the baseline algorithm, but rather that of the refined algorithm. Row 1 contains the refinement proposed in this section. Row 2 restates the baseline results. Row 3 is the baseline Furfine algorithm with the added restriction that the principal payment of an overnight loan is sent after 4 p.m. Row 4 is the additional requirement that the inferred interest rate must be rounded to the nearest basis or half-basis point, with a check that the repayment value is consistent with the rounded interest rate. Finally, row 5 is the same requirement as row 4 without the check on the repayment value. The last column indicates how restrictive the refinement is relative to the baseline algorithm as a proportion of loans identified on the original sample.

[12] While these results show that the proportional amount of false positives found by these refinements is higher than the baseline algorithm, the absolute amount of false positives is in fact lower.

5 Other Sources of Type I Error

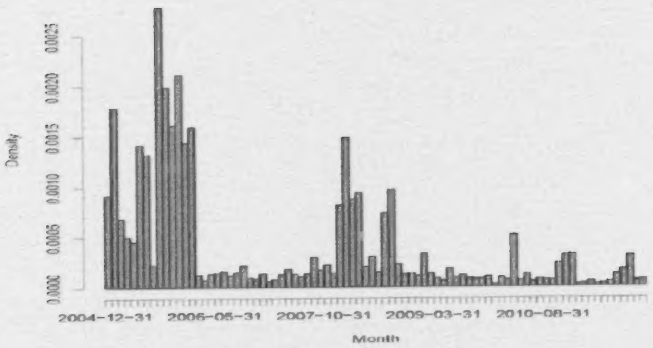
5.1 Foreign Exchange Transactions

The estimated bounds on the rate of false positives in the previous sections quantify the susceptibility of the Furfine algorithm and its variants to spurious matching. These bounds, however, do not cover other potential sources of misidentification. In particular, the permutation method does not fully capture the exposure of the algorithm to other non-loan payments that have interday dependence, such as foreign exchange transactions.

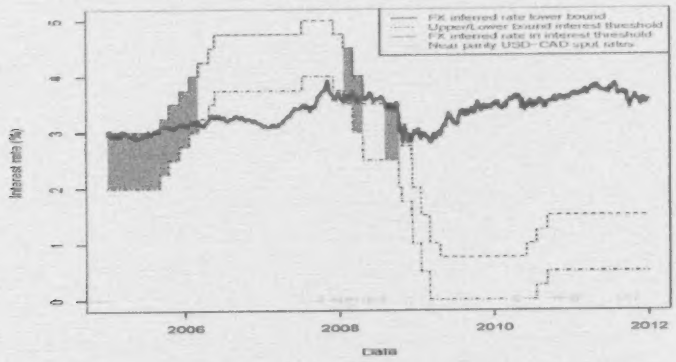
Foreign exchange transactions account for a substantial share of payments in RTGS systems. In Canada, the USD-CAD exchange market is by far the most dominant, and according to market conventions largely transacted in million-dollar increments, or its exchange rate equivalent quoted to the fourth decimal place, otherwise known as pip (Hendry and Kamhi (2009)).[13] Computing the CAD equivalents of round US$1 million payments (using the unique set of CAD-USD exchange rates quoted at the close of every minute within a given day[14]) and merging these simulated flows with the LVTS transactions yields an estimated 1.64% of the sample that is commensurate with an FX transaction (hereafter termed 'FX-consistent').

Figure 3a depicts the changing proportion of FX-consistent transactions in the false discoveries by month. The high rates observed between 2005Q1 and 2006Q1, in conjunction with the outlying levels of false discovery observed in this segment of the sample (see row 1 of Table 1), suggests a non-stationary relationship between the FX market and misidentification of overnight loans. There are two main ways for the Furfine algorithm to falsely identify a pair of independent FX transactions as an interbank loan:[15] either (i) the pair of FX transactions are both conversions from USD to CAD, with the difference in spot rates (annualized) satisfying the interest rate threshold; or (ii) a USD-CAD transaction is paired with a CAD-USD transaction, with the difference in spot rate from parity satisfying the interest rate threshold. Using the market convention that exchange transactions are quoted to the nearest hundredth of a cent, the lower bound on the inferred interest rate between two USD to CAD transactions (involving the same USD value on both days) is $\frac{0.0001}{SpotRate} \times 365 \times 100$.[16]

[13] Hendry and Kamhi (2009) note that the average FX transaction in the CAD-USD market was US$2 million.
[14] Data from FOREXITE.com.
[15] As opposed to FX swaps, wherein both transactions are part of the same trade.
[16] With repay = SpotRate' x Value and pay = SpotRate x value, the implied Effective Rate is then $\frac{repay-pay}{pay} \times 365 \times 100 = (\frac{SpotRate'}{SpotRate} - 1) \times 365 \times 100 = \frac{SpotRate'-SpotRate}{SpotRate} \times 365 \times 100$. The spot rate used is the maximum minute closing rate offered in the day (obtained from FOREXITE.com).



(a) Distribution of the number of transactions from the identified type I error pairs consistent with round million USD to CAD transfers by month, 2005Q1 - 2011Q4



(b) Path of Foreign Exchange Spot Rates and Interest Rate Thresholds
The crosses depict days where the max/min daily USD-CAD spot rate was sufficiently close to parity to yield an interest rate of 5% or less (e.g., 1 million /CAD - $(1/SpotRate) \times 1$ million USD ≈ 0). The shaded area is where the inferred lower-bound rate falls within the interest rate thresholds of the Furfine algorithm.

Figure 3: Interaction of Foreign Exchange Spot Rates and Interest Rate Filters

The path of this FX bound is given in blue in Figure 3b, with the dotted lines depicting the interest rate thresholds used for the baseline algorithm (overnight target $\pm$ 50 basis points). While in theory the FX bound below the upper bound of the interest rate threshold should be a necessary condition, this same bound simultaneously falling above the lower bound (shaded area in Figure 3b) corresponds almost perfectly with the spikes in 'FX-consistent' transactions in the false discoveries in Figure 3a. This correspondence suggests that much of the spurious matches in the sample are generated by these USD-CAD conversions and single pip differentials. The alternative source of potential pairs of FX transactions, depicted by the crosses in Figure 3b satisfying the interest rate threshold, appears to have a negligible effect in this sample, with only twenty days satisfying this requirement.

These results, combined with the introduction of intraday CLS settlement of FX transactions in many RTGS systems (the LVTS adopted such a settlement in 2013), imply that the algorithm's exposure to type I error due to FX is currently much less of a concern. Nonetheless, for historical samples it is important to control for the FX exposure. This can be done by either adjusting the admissible interest rate corridor for the algorithm such that this inferred FX bound lies outside of the corridor, or eliminating all FX-consistent transactions when the bound falls within this corridor.

6 Benefits of the Refined Algorithm

6.1 Bias Reduction

Improved accuracy of loan identification for the interbank lending market is important so far as it reduces bias in inferences. While inferences involving the number of transactions should be improved by the proposed k-NN refinement, inferences on other features such as the volumes and rates are not necessarily improved.

Comparing the implied distribution of interest rates (measured in deviations from the overnight target) of the refined algorithm against the marginal set of loans identified by the baseline Furfine algorithm but not by the k-NN refined variation, Figure 4a suggests an upward bias in the estimated rates arising from the baseline algorithm. Further, the deviance of the baseline Furfine rates from the overnight target displays significantly greater dispersion, as measured by the Bartlett test for equal variances at any reasonable confidence interval ($K^2 = 3090.98$).

This implied bias persists for the value-weighted average rate (referred to as the effective federal funds rate in the United States) used for policy-making, where using the Wilcoxon signed rank test an upward bias is detected at the 99% significance level. Although the median bias is small at 0.02 of a basis point, on 5.6% of the days the implied effective rates differ by over half a basis point, which can be the difference between intervention and non-intervention.

As Afonso, Kovner, and Schoar (2011) demonstrated, frictions in the interbank lending market are not necessarily reflected in prices, but sometimes, as occurred during the failure of Lehman Brothers, they are reflected in the quantities. As could be expected, the implied upward bias in the daily estimated volumes from the baseline Furfine algorithm is substantial. Figure 4b depicts the upward bias in the volumes, with an average difference of $234 million with 15 days in-sample exceeding differences of $1 billion.

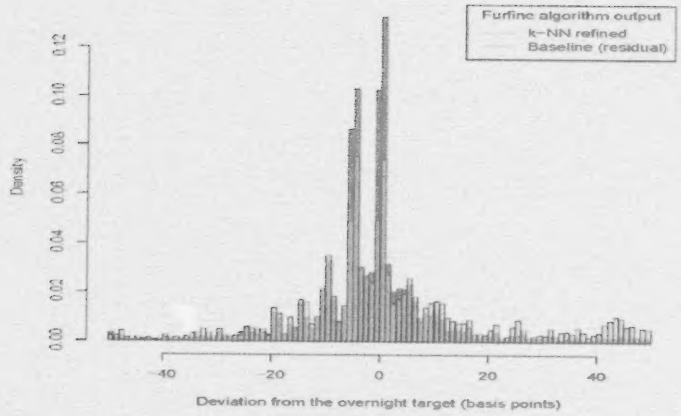
The evolution of loan volumes and quantities within the day is important because bargaining positions change as the search time horizon shortens. For instance, Ashcraft and Duffie (2007) find that a substantial number of loans are made earlier in the day by lenders who themselves are short relative to their reserves. Figure 4c depicts a bias in the number and volume of overnight loans toward earlier in the day by the baseline algorithm, suggesting that the extent of loans and positions may be substantially biased earlier in the day.

6.2 High-Frequency Tracking and Policy-Making

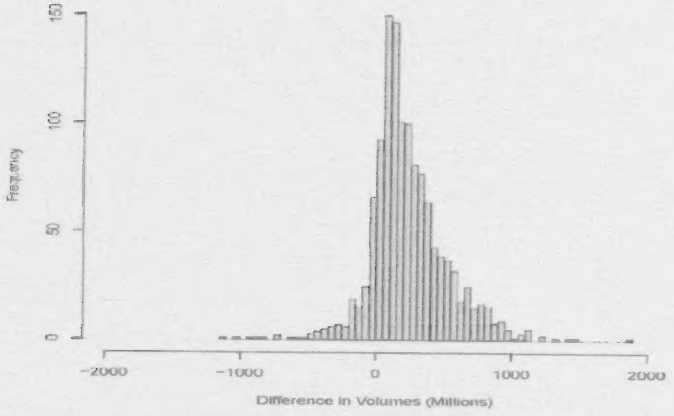
The Bank of Canada closely monitors the interbank market to maintain its target interest rate. Observing a significant deviation in the prevailing rates from their target, the Bank of Canada will offer Government of Canada securities for cash, or vice-versa, at the target rates.[17] The information which predicates a decision to intervene is itself based on direct surveys of the primary dealers and observed interdealer repo trades conducted in the CDSX, whose end-of-day value-weighted average reported by the interdealer brokers results in the Canadian overnight repo rate average (CORRA).

This information may not be representative of the overnight market for a number of reasons. First, the observed volumes may not be sufficiently large to accurately estimate the prevailing effective overnight rate. Second, there may be systematic differences between the interdealer brokered overnight repo market and the broader, bilateral repo market, between the repo and the

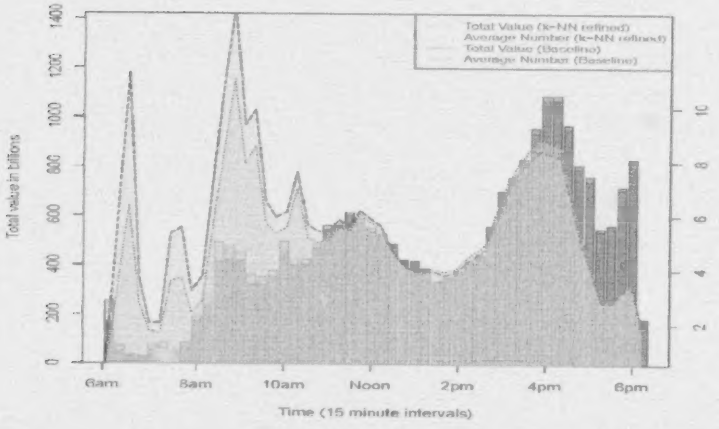
[17]For greater institutional detail on the specific means by which these interventions are implemented in Canada, see Reid (2007).



(a) Bias in Interest Rate Distribution



(b) Bias in Daily Loan Volumes



(c) Bias in Intraday Loan Volumes and Quantities

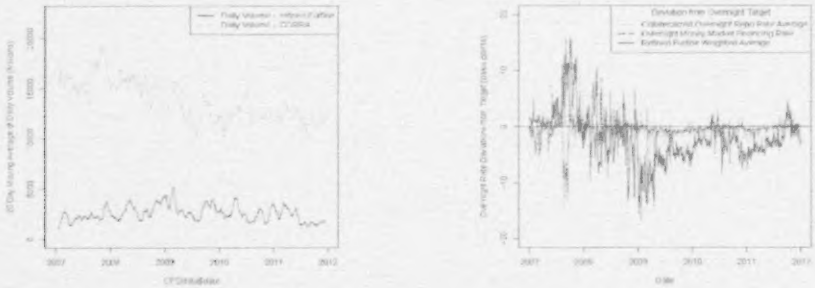
Figure 4: Implied bias of the baseline Furfine algorithm relative to k-NN refinement (Baseline - k-NN refined algorithm), January 2005 - December 2011.

uncollateralized loan market, or between the primary dealers and participants in the secondary market.

The quantity of loans used to compute the daily CORRA rate is quite small. Historically, the CORRA has been computed on the basis of only 10-15 transactions a day. Further, the aggregate daily volume for the computation of the CORRA only accounts for approximately 5% of the estimated daily volume of Government of Canada repos (Reid (2007)).[18]

Related to the CORRA (with an in-sample correlation of 84%) is the overnight money market financing rate, which is computed from a broader survey of dealers on the volumes and rates offered in the repo market. Its breadth comes at the cost of timeliness, since the survey occurs only once a day, as opposed to the real-time tracking of the on-screen interdealer brokered trades for the CORRA.

In contrast to these two measures, the refined Furfine algorithm is sampling from a broader pool of the overnight market. The inferred difference in market size is shown in Figure 5a. The top line shows the 20-day weighted average daily volume of overnight loans as measured by the refined Furfine algorithm, which is at least twice as large as the volumes from the CORRA (bottom line). Thus, even after factoring out an overly conservative 12.5% as false positives, the measured levels account for a far larger proportion of the estimated $55 billion overnight transactions.



(a) Overnight Loan Volumes **(b)** Effective Overnight Loan Rates

Figure 5: Comparison of implied rates and quantities from refined Furfine algorithm against leading alternatives.
Figure 5a compares the volumes recorded by the sources for the CORRA, the interdealer brokers, between 1 January 2007 and 31 December 2011. Figure 5b depicts the value-weighted mean spreads to the Bank of Canada overnight target, comparing the output from the kNN-refined Furfine algorithm against two measures of the collateralized overnight market, the CORRA and OMMF.

[18] According to quarterly trading statistics provided to the Bank and the Investment Dealers Association, the daily volume of Government of Canada repos was approximately $55 billion, as compared to the daily average interdealer volume of approximately $2.6 billion between 2007 and 2011.

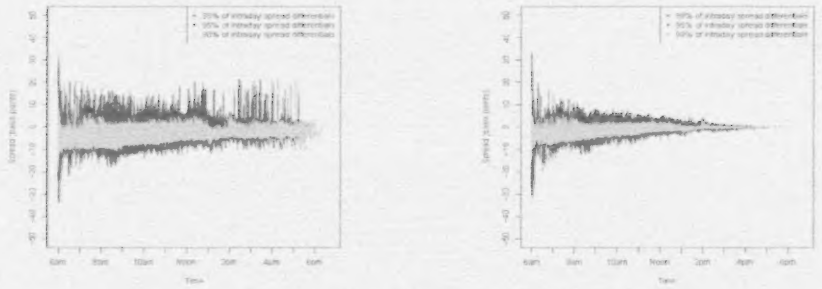
Although the CORRA, OMMF and the effective rate inferred from the refined Furfine algorithm are interrelated (correlation coefficient of 99.99% between CORRA and the effective Furfine rate), Figure 5b suggests that the markets sampled by each measure are distinct, with a correlation coefficient between the CORRA and inferred effective rate of -0.121. This negative correlation reflects the substitutability of the repo and uncollateralized market, especially in times of stress, which was particularly pronounced in the repo market near-collapse contemporaneous with BNP Paribas' asset freeze (see Figure 5b).

With minor modifications, the k-NN refined Furfine algorithm could be applied on a high-frequency basis to monitor the overnight market. Each time a payment appears on a given day, the algorithm could run in the reverse direction, looking for a matching principal payment the day before and updating the measured effective overnight rate as the return legs of overnight loans are identified. While they are a lagging indicator of the overnight conditions due to the identification of loans contracted the previous day, the data can be useful to identify frictions in the uncollateralized market that may persist across several days (as is often the case for tightness in the repo market).

The minute-by-minute spread of the effective rate against the end-of-day CORRA computed in such an aforementioned manner is shown in Figure 6a for the post-effective lower-bound period. Here we see that, 95% of the time, the relationship between the spread and the effective Furfine rate falls within 10 basis points of each other and converges to a tighter relationship by day's end. The convergence of the intraday effective Furfine rate to the end-of-day effective Furfine rate is quite fast, being within 7 basis points of the end-of-day rate with 2 hours of operation, within 4 basis points by noon and 1 basis point by 3 p.m. (similar dynamics exist for the other periods, though the crisis and ELB periods intuitively show higher variance and slower convergence).

7 Conclusion

Transaction-level data on the interbank lending market offers a unique window into the functioning of OTC markets in general. The lack of formal evidence on the accuracy of the Furfine (and variant) algorithm's identification of overnight loans has cast doubt on the usefulness of the data it provides. In this paper, I estimate the daily rate of false positives associated with the algorithm through a permutation-style approach. Applied to the LVTS between January 2005



(a) Intraday Furfine-CORRA Spread, June 2010 - December 2011.

(b) Intraday Furfine Rate Convergence, June 2010 - December 2011.

Figure 6: Intraday convergence of refined Furfine inferred effective rate.
Figure 6a compares the effective overnight rate computed up to the minute (based on the return leg of an identified loan) against the end-of-day CORRA. Figure 6b depicts the convergence of the up-to-the-minute effective rate against its final end-of-day value (weighted average of all overnight loans returned that day).

and December 2011, I demonstrate that the false positive rate of the algorithm is not bounded below 10% with 95% confidence.

I propose a refinement of this algorithm that, in contrast to the baseline Furfine algorithm and other standard variants, bounds the false positive rate below the 10% level (with 99% confidence). This refinement is an improved treatment of the situation where a payment is identified as a leg in two or more overnight loans. Unlike other proposed treatments in the literature that only choose amongst the payment pairs, my treatment reduces the algorithm's exposure to a type I error. I do this by comparing the inferred covariates of the uniquely identified overnight loans and randomly paired transactions against these duplicate loans. I also assess the exposure of the Furfine algorithm to foreign exchange flows and propose a heuristic means to mitigate this form of type I error based on the interaction of the inverse spot rate and the upper/lower interest rate bounds of the algorithm.

Importantly, both the testing and refinement methods used here can easily be applied to another nation's RTGS system. Of course, external data on the identity of overnight loans may further enhance the performance of the machine-learning methods applied, and an assessment of the other sources of type I error. Finally, while this paper only assesses the identification of interbank overnight loans, future work could ascertain the associated levels of accuracy for term interbank loans (Kuo et al. (2013)).

Although the estimated false positive rates are only upper bounds on the amount of type I error associated with spurious matching and may not capture all sources of type I error, par-

ticularly misidentification at the counterparty level, the methodology should inform interested parties about the degree of sensitivity of any resulting inferences from this data. As displayed by the estimation approach in this paper, the refinements proposed provide a new tool to extract interbank loans from payment data with confidence that the type I error is small.

Bibliography

Acharya, Viral V and Ouarda Merrouche. 2013. "Precautionary hoarding of liquidity and interbank markets: evidence from the subprime crisis." *Review of Finance* 17 (1):107–160.

Afonso, Gara, Anna Kovner, and Antoinette Schoar. 2011. "Stressed, not frozen: The federal funds market in the financial crisis." *Journal of Finance* 66 (4):1109–1139.

Allen, Jason, Ali Hortaçsu, and Jakub Kastl. 2011. "Analyzing default risk and liquidity demand during a financial crisis: The case of Canada." *Bank of Canada Working Paper* No. 2011-17.

Arjani, Neville and Darcey McVanel. 2006. "A Primer on Canada's Large Value Transfer System." Tech. rep., Bank of Canada.

Armantier, Olivier and Adam Copeland. 2012. "Assessing the Quality of 'Furfine-Based' Algorithms." *FRB of New York Staff Report* No. 575.

Ashcraft, Adam B and Darrell Duffie. 2007. "Systemic illiquidity in the federal funds market." *American Economic Review* 97 (2):221–225.

Chu, Ba, Chufeng Huang, Kim P. Huynh, and Lanh Tran. 2013. "Adaptive Permutation Tests for Serial Independence." *Statistica Neerlandica* (forthcoming).

Demiralp, Selva, Brian Preslopsky, and William Whitesell. 2006. "Overnight interbank loan markets." *Journal of Economics and Business* 58 (1):67–83.

Efron, Bradley. 2004. "Large-scale simultaneous hypothesis testing." *Journal of the American Statistical Association* 99 (465).

Furfine, Craig H. 1999. "The Microstructure of the Federal Funds Market." *Financial Markets, Institutions & Instruments* 8 (5):24–44. URL `http://dx.doi.org/10.1111/1468-0416.00031`.

Hastie, Trevor, Robert Tibshirani, and Jerome H. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Berlin: Springer-Verlag, second ed.

Hendry, Scott and Nadja Kamhi. 2009. "Uncollateralized overnight lending in Canada." *Applied Financial Economics* 19 (11):869–880.

Kuo, Dennis, David Skeie, James Vickery, and Thomas Youle. 2013. "Identifying term interbank loans from Fedwire payments data." *Federal Reserve Bank of New York, Working Paper* No. 603.

Ploberger, Werner and Walter Krämer. 1992. "The CUSUM test with OLS residuals." *Econometrica: Journal of the Econometric Society* 60 (2):271–285.

Reid, Christopher. 2007. "The Canadian overnight market: Recent evolution and structural changes." *Bank of Canada Review* (Spring):15–29.

Storey, John D. 2002. "A direct approach to false discovery rates." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (3):479–498.

Storey, John D and Robert Tibshirani. 2001. "Estimating the positive false discovery rate under dependence, with applications to DNA microarrays." URL `http://genomics.princeton.edu/storeylab/papers/dep.pdf`.

Zeileis, Achim. 2000. "Alternative boundaries for CUSUM tests." *Statistical Papers* 45 (1):123–131.